Early screening of autism spectrum disorder using cry features

Aida Khozaei[1], Hadi Moradi[1, 3*], Reshad Hosseini[1], Hamidreza Pouretemad[2], BaharehEskandari[2]

[1] School of Electrical and Computer Engineering, University of Tehran, Tehran, Iran.

[2] Department of Psychology, Shahid Beheshti University, Tehran, Iran

[3]Adjunct Research Professor, Intelligent Systems Research Institute, SKKU, Suwon, South Korea

[*]Corresponding author

E-mail: moradih@ut.ac.ir (HM)

Abstract

Due to the importance of automatic and early autism screening, in this paper, a cry-based screening approach for children with Autism Spectrum Disorder (ASD) is introduced. During the study, we realized that the ASD specific features are not necessarily observable among all children with ASD and among all instances of each child. Therefore, we proposed a new classification approach to be able to find such features and their corresponding instances. We tested the proposed approach and found two features that can be used to distinguish groups of children with ASD from Typically Developing (TD) children. In other words, these features are present in subsets of children with ASD not all of them. The approach has been tested on a dataset including 14 boys and 7 girls with ASD and 14 TD boys and 7 TD girls, between 18 to 53 months old. The sensitivity, specificity, and precision of the proposed approach for boys were 85.71%, 100%, and 92.85%, respectively. These measures were 71.42%, 100%, and 85.71% for girls, respectively.

Keywords Screening, Autism spectrum disorder, Crying, Clustering, Classification, Feature Selection

Introduction

Children with Autism Spectrum Disorder (ASD) are defined by their abnormal or impaired development in social interaction and communication, as well as restricted and repetitive behaviors, interests, or activities [1]. The rapid growth of ASD in the past 20 years has inspired many research efforts toward the diagnosis and rehabilitation of ASD [2-5]. In the field of diagnosis, there are several well-established manual methods to diagnose children over 18 months [6]. However, the practical average age of diagnosis is over 3 years due to the lack of knowledge about ASD and the lack of widespread expertness for autism diagnosis [7, 8]. Considering the fact that early diagnosis is crucial for effective treatments [7, 9], there are two main questions: 1) can autism be diagnosed earlier than 18 months and 2) is it possible to employ intelligent methods for screening of autism to eliminate the widespread need for experts?

Fortunately, there are studies showing that the age of screening can become lower than 18 months. For example, Thabtah and Peebles [10] reviewed several questionnaire-based approaches that may be able to screen ASD above 6 months of age. However, those approaches, like Autism Diagnostic Interview-Revised (ADI-R) [11] and Autism Diagnostic Observation Schedule (ADOS) [12] which have been clinically proven to be effective and adequate, are time-consuming instruments [10] and need trained practitioners to use them. To reduce the dependency on the human expertise needed in using such questionnaires [8], several studies proposed machine learning methods to classify children with ASD [13, 14] using questionnaires. Their goal is to make the process automatic and/or find an optimum subset of questions or features. For instance, Abbas et al. [15] proposed a multi-modular assessment system combined of three modules, a parent questionnaire, a clinician questionnaire, and a video assessment module. Although the authors used machine

learning to automate and improve classification process, however, it still needs human involvement to answer questions or assess videos.

As mentioned above, there are studies tried to screen children with ASD under 18 months, however, they still need expertise and/or need manual processing. On the other hand, Emerson et. al showed that fMRI [16] can be used to predict the diagnosis of autism at 2 years of age of high-risk 6-month-old infants. Denisova and Zhao [17] used movement data from rs-fMRI from 1-2 month-old infants to predict future atypical developmental trajectories as biological features. Furthermore, Bosl, Tager-Flusberg, and Nelson [18] suggested that useful biomarkers can be extracted from EEG signals for early detection of autism. Blood-based markers [19, 20] and prenatal immune markers [21] were also proposed to diagnose ASD that can be used right after birth. Although these approaches suggest new directions toward early ASD diagnosis, however, these are costly, need expertness, and need dedicated equipment, which would limit their usage. Furthermore, these are still at the early stages of research and need further approval. Finally, approaches which involve methods such as fMRI or EEG, are hard to be used on children, especially on children with autism who may have trouble following instructions appropriately [22], have atypical behaviors [23], or have excessive head movements [24, 25].

There are studies that used vocalization-based analysis to screen children with autism. For instance, Brisson et al. [26] showed voice features' differences between children with ASD and Typically Developing (TD) children. Several studies, like [27], used speech-related features for the screening of children older than 2. To reach the goal of early ASD screening, vocalizations of infants under 2 years of age have been investigated [28-30]. Santos et al. [28] used vocalization, such as babbling, to screen ASD children at the age of 18 months. They collected data from 23 and 20 ASD and TD children, respectively. They reported high accuracy of around 97% which can

be due to the fact that they used k-fold cross-validation without considering subject-wise hold out to have unseen subjects in the test fold [31]. Oller et al. [29] proposed another vocalization-based classification method in which they included age and excluded crying. They applied the method on 106 TD children and 77 children with ASD between 16 to 48 months and reached 86% accuracy. Pokorny et al. [30] extracted eGeMAPS parameter set [32], which includes 88 acoustic parameters, on 10 months old children. This set consists of statistics calculated for 25 frequency-related, energy-related, and spectral low-level descriptors. They reached 75% accuracy on a population of 10 TD children and 10 children with ASD.

Esposito, Hiroi, and Scattoni [33] showed that cry is a promising biomarker for the screening of ASD children. Sheinkopf et al. [34] and Orlandi et al. [35] have shown that there are differences in the cry of children with ASD compared to TD children. To the best of our knowledge, our own group's preliminary study [36] was the only research that used cry voices for the screening of children with ASD. We used a dataset of 5 children with ASD and 4 TD children older than two years of age. The accuracy of the proposed method is 96.17% using k-fold cross validation without considering subject-wise hold out, which is a shortcoming of this study. In other words, it has been overfitted to the available data and may fail to correctly classify new samples. So, a thorough examination using an unseen test set on cry features is necessary to evaluate the results. It should be noted that the data from our previous study [36] could not be used in the study presented in this paper due to the differences in data collection procedures.

In all the above studies, it was assumed that the specific sound features, distinguishing children with ASD from TD children, are common among all the ASD cases. However, this may not be true for all the features. For instance, tiptoe walking, which is one of the repetitive behaviors of children with ASD, appears in approximately 25% of these children [37]. Consequently, in this

paper, we propose a new cry-based approach for screening children with ASD. Our screening approach makes use of the assumption that all discriminative characteristics of autism may not appear in all ASD children. This assumption is in contrast with the assumption in the ordinary instance-based machine learning methods, which assumes that all instances of a class include all discriminative features needed for classification. In our proposed method, first discriminable instances of the cry, which exist in subsets of children with ASD, are found. Then it uses these instances to extract and select features to distinguish between these ASD instances from TD instances. It should be mentioned that the final selected features, in this study, are common among our set of children with ASD between 18 to 53 months of age. These selected features support experiential knowledge of our experts stating that the variations in the cry of children with ASD are more than TD children. This approach is different from the other approaches that either used a dataset of children with a specific age [28, 30] or used age information for classification [29]. The proposed approach has been implemented and tested on 62 participants. The results show the effectiveness of the approach considering accuracy, sensitivity, and specificity of screening.

Method

Since this study was performed on human participants, first, it was approved by the ethics committee at Shahid Beheshti University of Medical Sciences and Health Services. All the parents of the participants were informed about the study and signed an agreement form to be included in the study.

Participants

There were 62 participants between 18 and 53 months that consisted of two groups, i.e. 31 ASD and 31 TD with 24 boys and 7 girls in each group. Since we expected to have different vocalization characteristics for boys and girls, the training set was assembled of only boys, including10 TD, and 10 ASD. In other words, we wanted to eliminate the gender effects on the feature extraction and model training. Unfortunately, due to the lower number of girls with ASD in the real world, not enough data for girls with ASD could be collected. Nonetheless, the model was also tested on the girls to see how it would generalize even on girls.

The inclusion criteria of the ASD participants were: a) had been just diagnosed with ASD based on DSM-5 with no other neurodevelopmental, mental, and intellectual disorder, b) there were no other known medical, genetic conditions, or environmental factors, and c) had not taken any treatment or medication, or had taken treatment less than a month. There were only two girls who did not fall into these criteria since they had been diagnosed for more than a year. The participants' average language development assessed based on [38-41] was between 6 to 12 months at the time of participation. The autism diagnosis procedure started with the Gilliam Autism Rating Scale-Second Edition (GARS-2) questionnaire [42] which was answered by the parents. Then the parents were interviewed, based on DSM-5, while the participants were evaluated and observed by two Ph.D. degree child clinical psychologists. In addition, the diagnosis of ASD was separately confirmed by at least a child psychiatrist in a different setting. TD children were selected from those in an age range similar to the ASD participants from volunteer families at homes and health centers. They had no evidence and official diagnosis of neurological or psychological disorder at the time of recording their voices. The children with ASD were older than 20 months with the mean, standard deviation, and range of 35.6, 8.8, and 33 months respectively. The TD children

were younger than 51 months with the mean, standard deviation, and range of about 30.8, 10.3, and 33 months respectively. It should be mentioned that the diagnosis of the children under 3 years was mainly based on experts' evaluation, not the GARS score. Furthermore, all TD participants under 3 years of age were followed up when they passed 3, to make sure the initial TD assignment was correct or still valid.. To do so, we used a set of expert-selected questions based on [43] to assess them through interviews with parents.

Tables 1 and 2 show the details of the training and the test sets' participants, respectively. In each table, the number of voice instances from each participant and the total duration of all its instances in seconds are shown in columns 3 and 4, respectively. The recording device category, i.e. a high-quality recorder (HQR) and typical cell phones (CP), is given in the device category column. The next two columns include GARS-2 scores and the language developmental milestone of the participants with ASD at the time of recording. In six cases, there were no GRAS scores available at the time of study (No Data (ND)). The column labeled as "Place" shows the location of the recording which can be homes (H), autism centers (C1, C2, and C3), and health centers (C4, C5, and C6). There were a total number of 359 samples for all children. 53.44 % of the samples were from ASD participants and 46.56 % were from TD participants.

Table 1.The training set participants' data

	ID	Age (month)	# of instances	Total duration (sec)	Device	GARS score	Language milestone (month)	Place	Reason of cry
ASD	ASD1	20	9	7.8	CP	104	0-6	C1	Annoyed/Uncomfortable
	ASD2	24	3	1.5	HQR	83	0-6	C2	Unwilling
	ASD3	26	5	2.1	HQR	120	0-6	C1	Annoyed/Uncomfortable
	ASD4	28	13	9.1	HQR	121	0-6	C2	Annoyed/Uncomfortable
	ASD5	29	14	26	HQR	89	6-12	C2	Unwilling/Complaining
	ASD6	31	4	2.4	HQR	87	0-6	C2	Unwilling/Complaining
	ASD7	36	11	11	HQR	87	6-12	C2	Unwilling/Complaining
	ASD8	43	2	0.7	CP	ND	ND	C2	Unwilling
	ASD9	45	3	2.6	CP	72	6-12	C2	Complaining

	ASD10	45	4	3.4	CP	ND	ND	H	Sleepy
TD	TD1	21	11	14	HQR	NA	NA	H	Complaining
	TD2	24	12	12	HQR	NA	NA	C4	Scared/Unwilling
	TD3	26	2	2.3	HQR	NA	NA	C5	Unwilling
	TD4	28	6	13	CP	NA	NA	C5	Scared/Unwilling
	TD5	36	3	2.6	CP	NA	NA	H	Unwilling/Complaining
	TD6	38	3	1.5	HQR	NA	NA	C6	Complaining
	TD7	41	3	2.4	HQR	NA	NA	H	Unwilling
	TD8	43	3	2.2	CP	NA	NA	H	Sleepy
	TD9	44	2	1.2	CP	NA	NA	H	Complaining
	TD10	51	2	1.7	CP	NA	NA	H	Complaining

Table 2.The test set information

		ID	Age (month)	# of instances	Total duration (S)	Device	GARS score	Language milestone (months)	Place	Reason of cry
ASD	Boys	ASD11	28	12	7.2	HQR	102	0-6	C2	Unwilling/ Uncomfortable
		ASD12	30	18	17.1	HQR	ND	ND	C3	Mother Separation
		ASD13	30	3	2.9	CP	ND	ND	H	Unwilling/Sleepy
		ASD14	31	5	2.3	HQR	73	0-6	C2/ H	Mother Separation/Hungriness
		ASD15	33	3	2.5	HQR	91	0-6	C2	Unwilling
		ASD16	33	2	2.5	HQR	104	0-6	C1	Annoyed/Uncomfortable
		ASD17	34	1	0.6	HQR	91	0-6	C2	Unwilling/Complaining
		ASD18	35	2	1.7	HQR	81	ND	C1	Annoyed/Uncomfortable
		ASD19	37	1	0.6	HQR	94	12-18	C2	Unwilling/Complaining
		ASD20	40	19	14	HQR	91	0-6	C1	Annoyed
		ASD21	45	1	0.3	HQR	81	6-12	C2	Unwilling/Complaining
		ASD22	48	2	1.6	HQR	100	6-12	C2	Annoyed/Complaining
		ASD23	52	6	3.1	HQR	113	12-18	C2	Unwilling/Complaining
		ASD24	53	7	5.2	HQR	78	6-12	C1	Annoyed/Uncomfortable
	Girls	ASD25	25	12	14	HQR	85	0-6	C2	Unwilling/Complaining
		ASD26	26	5	2	CP	102	0-6	C1	Scared
		ASD27	31	3	1.7	HQR	94	0-6	C2	Unwilling/Complaining
		ASD28	32	2	1.3	HQR	100	0-6	C2	Unwilling/Complaining
		ASD29	41	8	3	HQR	102	0-6	C2	Unwilling/Complaining
		ASD30	45	2	1.2	CP	ND	ND	H	Thirsty
		ASD31	49	7	12	CP	ND	ND	H	Unwilling/Complaining
TD	Boys	TD11	18	4	2	HQR	NA	NA	C4	Scared
		TD12	18	7	5.1	HQR	NA	NA	C4	Scared/Unwilling
		TD13	19	7	4.2	HQR	NA	NA	C5	Unwilling
		TD14	20	9	8	HQR	NA	NA	C5	Unwilling/Complaining
		TD15	21	4	1.2	HQR	NA	NA	H	Complaining
		TD16	24	3	2.7	HQR	NA	NA	C5	Scared /Unwilling
		TD17	24	2	1.5	HQR	NA	NA	C5	Scared/Unwilling
		TD18	24	6	5.1	HQR	NA	NA	C4	Unwilling/Complaining
		TD19	24	4	2.4	HQR	NA	NA	C5	Unwilling/Complaining
		TD20	24	5	4.2	HQR	NA	NA	C5	Unwilling/Complaining
		TD21	29	11	10	HQR	NA	NA	H	Unwilling/Complaining
		TD22	30	4	2	HQR	NA	NA	C5	Scared/Unwilling

		TD23	30	4	2	CP	NA	NA	H	Unwilling
		TD24	43	12	11	HQR	NA	NA	H	Complaining
	Girls	TD25	24	5	6	HQR	NA	NA	C4	Unwilling/Complaining
		TD26	25	2	4.4	HQR	NA	NA	C5	Scared
		TD27	29	5	5	HQR	NA	NA	C5	Scared
		TD28	33	2	2.1	CP	NA	NA	H	Complaining
		TD29	45	16	11	HQR	NA	NA	H	Unwilling/Complaining
		TD30	50	6	7	HQR	NA	NA	H	Complaining
		TD31	51	2	0.7	CP	NA	NA	H	Unwilling

Two groups of 10 TD and 10 ASD children were selected for training the classifiers such that two groups were as balanced as possible with respect to the age and the recording device. Thus, each child in the TD group had a corresponding child in the ASD group with about the same age. As a result of this balancing of our data, we obtained training participants with the age between 20 and 51 months. The mean ages in the training set were 32.7 and 35.2 months for ASD and TD participants, respectively. The standard deviations are 9 and 9.9 months with the range of 25 and 30 months for ASD and TD participants, respectively.

Although this approach was trained and tested on children older than 18 months, we tested the approach on 57 participants between 10 to 18 months to investigate how it works on children under 18 months. These 57 participants consisted of 28 boys and 29 girls with the mean of 15.2 for both and standard deviations of 2.8 and 2.9 respectively. All these participants were evaluated afterward at the age of 3 or more by the same follow-up procedure using our expert-selected questionnaire.. At the time of initial voice collection, 55 of these participants had no evident or diagnosed disorder. Two of them were referred to our experts due to the positive results of screening by our method. The diagnosis or concerns for the two mentioned participants as well as the participants with any evidence of having abnormality in the developmental milestones during the follow-up procedure are summarized in Table 3. The summary of disorders is given in the last column of Table 3 is based on the parental interviews and our experts' evaluation. Unfortunately, Child5, Child6, and Child7's parents did not cooperate to get experts' evaluation.

Table 3. The participants with an abnormality in the follow-up

ID	Gender	Age (in months)		Disorder
		at recording time	at following-up time	
Child1	M	11	11	Developmental delay[a], signs of genetic diseases
Child2	M	17	17	UNDD[b]
Child3	M	12	40	ASD[b]
Child4	M	12	36	Sensory processing disorder[c], several ADHD symptoms[b]
Child5	M	18	40	Language delay
Child6	M	15	46	Developmental delay symptoms
Child7	M	12	43	Developmental delay symptoms

UNDD, Unspecified Neurodevelopmental Disorder.

[a] Clinical observation by our expert based on [43].

[b] Clinical observation by our expert based on [1]

[c] Clinical observation by our expert based on [44].

Data collection and preprocessing

As mentioned earlier, the data was recorded using high-quality devices and typical smartphones. The high-quality devices were a UX560 Sony voice recorder and a Sony UX512F voice recorder. To use typical smartphones, a voice-recording and archiving application was developed and used on various types of smartphones. All voices, through the application or the high-quality recorders, were recorded in wav format, 16 bits, and with the sampling rate of 44.1 kHz. The reason for using various devices was to avoid biasing the approach to a specific device. Similarly, the place of recording was not restricted to one place to make the results applicable to all places.

The parents and trained voice collectors were asked to record the voices in a quiet environment. Furthermore, they were asked to keep the recorders or smartphones about 25 cm from the

participants' mouth. Despite the proposed two recommendations, there were recorded voices without following the recommendations and did not have the required quality. Thus, those recordings were eliminated from the study. Also, all the cry voices that were due to pain had been removed from the study since they were similar between the TD and ASD groups.

After data collection, there was a preprocessing phase in which only pure cry parts of the recordings, with no other types of vocalization, were selected. To explain more, the parts of cry voices which were accompanying screaming, saying words/other vocalizations, or were occurred with closed/non-empty mouth were eliminated. All segmentations and eliminations were done manually using Sound Forge Pro 11.0. From the selected cries, the beginning and the end, which contain voice rises and fades, were removed to keep only the steady parts of the cries. It prevents having too much variation in the voice which can lead to unsuitable statistics. Also, the uvular/guttural parts of the cries were removed. The reason is that we believe these parts distort the feature values of the steady parts of a voice. Each remaining continuous segment of the cries was considered and used as a sample (instance) in this study. Finally, since the basic voice features were extracted from 20 milliseconds frames, to generate statistical features of the basic features, the minimum length of the cry segments were set to 15 frames, i.e. 300 milliseconds. Thus, any cry samples below 300 milliseconds were eliminated from the study. In this study, the final prepared samples were between 320 milliseconds to 3 seconds.

Feature extraction

Previous studies working on voice features for discriminating ASD children use different sets of features. These methods share several common features like F0, i.e. the fundamental frequency of a voice, and Mel-Frequency Cepstral Coefficients (MFCC), i.e. coefficients which represent the

short-term power spectrum of a sound [45]. F0 has been one of the most common features used [26, 27, 34]. However, since age is an important factor affecting F0 [46], this feature is useful when participants have a similar age. On the other hand, MFCC coefficients and several related statistical values have been reported to be useful features in several studies [30, 36, 47]. Considering useful features reported in the previous studies and the specifications of the current study, several features were selected to be used in this work that are explained in the following.

In this study, each instance was divided into 20 milliseconds frames, to extract basic voice features. We used several features proposed by Motlagh, Moradi, and Pouretemad [36] and by Belalcázar-Bolaños et al. [48]. The features used by Motlagh, Moradi, and Pouretemad [36] include certain statistics like mean and covariance of the frame-wise basic features, like MFCC coefficients, over a voice segment. They also used the mean and variance of frame-wise temporal derivative [49, 50] of the basic features. The frame-wise temporal derivative means the difference between two consecutive frames, which in a sense is the rate of change of a feature value in one frame step. We modified the spectral flatness features by including the range of 125-250 Hz beside the 250-500 Hz range. This range was added to cover a wider frequency range than the normal children frequency range, which showed to be necessary in the process of feature extraction and selection. Each range is divided into 4 octaves and the spectral flatness is computed for those octaves.

We removed all uninformative and noisy features of the set which are explained in the following. The mean of frame-wise temporal derivative of the basic features is removed because it is not a meaningful feature and is equal to taking the difference between the value of the last and the first frames. There are means of the features related to the energy, such as the audio power, total loudness, SONE, and the first coefficient of MFCC, that were removed to make the classifier

independent of the loudness/power in children's voices. Zero crossing rate (ZCR) is omitted too, due to its dependency on the noise in the environment.

The second set of features used in this study is from Belalcázar-Bolaños et al. [48] because it has phonation features, like jitter and shimmer. Jitter and shimmer, which have been reported to be discriminative for ASD, are linked to perceptions of breathiness, hoarseness, and roughness [51]. Other features used from Belalcázar-Bolaños et al. [48] include glottal features related to vocal quality and the closing velocity of the vocal folds [28]. The mean of logarithmic energy feature is omitted for the same reason as other energy-related features. A summary of the features added to or removed from the sets by [36] and [48], is presented in Table 4.

Table 4. The features and statistics which were added or removed to the two feature sets.

	Feature	removing/adding	Reason
Second set	logarithmic energy	Mean statistic is removed	Classification dependency on loudness/power of cry
First set	Audio power		
	Total loudness		
	SONE		
	First MFCC coefficient		
	ZCR	The basic feature is removed	The feature's dependency on environmental noise
	All basic features applicable	mean of frame-wise temporal derivative of the basic features is removed	No meaning for the feature
	MFCC	Coefficients of 14-24 are added	Having higher-order coefficients for vocal cords information as well as vocal tract
	Spectral flatness	A range of 125-250 Hz is added	Covering the low-frequency range of human voice

The proposed subset instance classifier

To explain the proposed classifier, let's assume that there is a target group of participants that we want to distinguish from the rest of the participants, called the rest. Furthermore, each participant in the target group may have several instances that may be used to distinguish the target group from the rest. Fig 1a shows a situation in which all instances of all participants of the target group are differentiable using common classifiers that we call Whole Set Instance (WSI) classifiers. In this figure, the circles represent our target group and the triangles represent the rest. The color coding is used to differentiate between the instances of each participant among each group. In contrast to the situation in Fig 1a, in Fig 1b the target group cannot be easily distinguished from the rest. In such a situation, there are instances of two participants in the target group, i.e. the red and brown circles that are not easily separable from the instances in the rest (Case 1). Furthermore, there is a participant with no instances, i.e. the orange circles, easily separable from the rest (Case 2). An example of Case 1 is tiptoe walking in children with ASD, which is common in about 25% of these children [37] who do it most of the time. An example of Case 2 is children with ASD who do not tiptoe walk. In other words, there are children with ASD who cannot be distinguished from TD children using the tiptoe walking behavior.

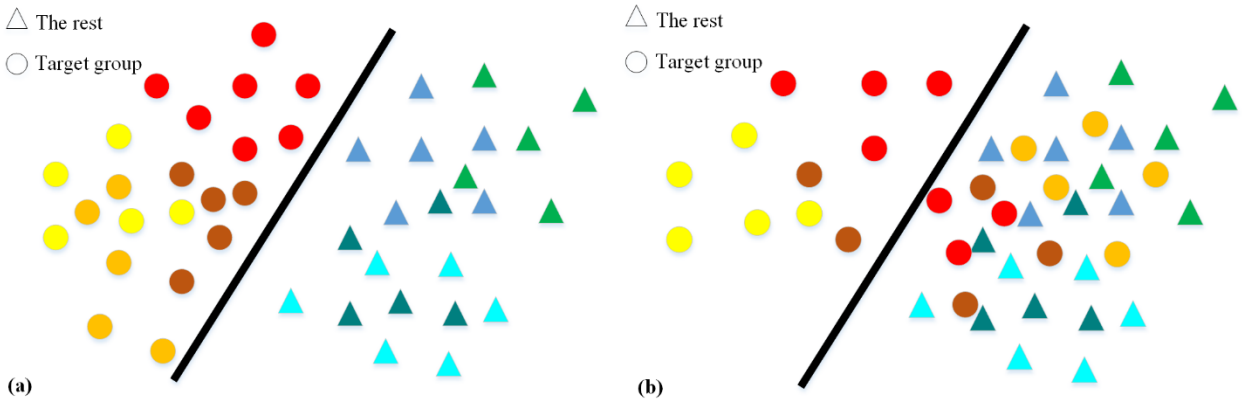


Fig 1. Two different hypothetical types of two-dimensional data of the target group and the rest. The instances shown by the warm-colored circles and the cool-colored triangles are for the target group and the rest, respectively. All instances belonging to a participant have the same color. In (a), all the target group participants' instances are distinguishable using a classifier. In (b), only some instances of the target group participants are separable from the other instances by a classifier.

Applying any WSI classifier may fail for the data type shown in Fig 1b. Consequently, we proposed SubSet Instance (SSI) classifier that first finds differentiable instances and then trains a classifier on these instances. As an example, the proposed SSI classifier first tries to find the circles on the left of the line in Fig 1b, using a clustering method. Then, it uses these circles, as *exclusive instances* having a specific feature common in a subset of the target group, to train a classifier separating a subset of the target group.

The steps of common WSI classifiers are shown in Fig 2a. The steps of our proposed SSI classifier are shown in Fig 2b. In the SSI classification approach, after the feature extraction and clustering steps, for each cluster, a classifier is trained to separate its exclusive instances from the instances of the rest of the participants. In the testing phase, any participant with only one instance classified

in the target group (positive instance), is classified as a target group's participant. The pseudo-code for the proposed approach is given in Algorithms 1 and 2.

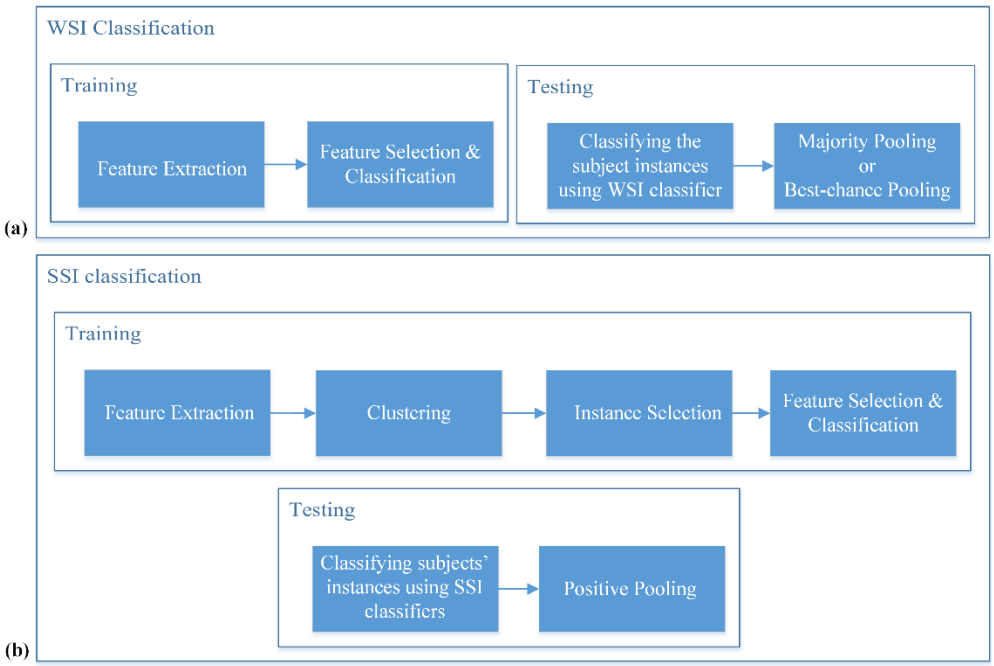


Fig 2. An overall view of WSI and SSI methods. (a) In WSI method, after feature extraction, a classifier is trained on all instances and majority pooling (MP) is usually used in the testing phase. In this study Best-chance threshold Pooling (BP), which is a threshold-based pooling with the threshold giving the best accuracy on the test set, is also used to give the best chance to WSI classifier. (b) In the proposed SSI classifier, after feature extraction, clustering is applied to find and select exclusive instances containing instances of the target group participants only. Then classifiers are trained using exclusive instances, and a participant is classified in the target group in the testing phase if any classifier detects a positive instance for it.

Algorithm 1. Training SSI classifiers

T: set of all target group instances

R: set of all the rest instances

F: set of all Classifiers

ρ: threshold for the number of samples in a cluster

s: the number of minimum samples needed in a cluster to be able to train a classifier for it

C_j: The jth cluster

n: number of clusters

$F = \emptyset$

1: While $\exists j \, |C_j| > \rho$; while there is a cluster bigger than a threshold or n =1

2: $n = n + 1$; increase the number of clusters

3: Cluster the $T + R$ into n clusters $C_j, j = 1, \ldots, n$

4: $EC = \{C_j \subset T\}$; the set of clusters of only exclusive instances, i.e. exclusive clusters

5: If $EC \neq \emptyset$; check if there is any exclusive cluster

6: For all C_j in EC with $|C_j| > s$

7: Train a classifier using positive labels c $\in C_j$ and negative labels $r \in$ R

8: Add the classifier to F

9: $T = T - \sum_{C_j \subset EC} C_j$; remove the instances of the exclusive clusters from target group instances

10: $n = 1$; set 1 to re-start clustering in two groups on the remaining instances

Algorithm 2. Testing SSI classifiers

F: set of trained classifiers

A: set of subject instances

1: For all instances a of A

2: $P = \{a \in A | \exists f, classifies\ a\ as\ positive\ instance\}$

3: If $P \neq \emptyset$

4: The participant is from the target group

5: Else

6: The participant is from the rest

In the proposed training algorithm of the SSI approach, the goal is to find clusters containing the ASD instances only. Then a classifier is trained using the instances of these clusters and added to a list of all trained classifiers (lines 7 and 8 of Algorithm 1). As shown in the loop of the algorithm, starting at line 1, the data is clustered starting with two clusters. Then the number of clusters is

increased until a cluster containing only the target group instances emerges. The exclusive instances in such a cluster are removed from the set of all target group's instances, and the loop is restarted. Before restarting the loop if the number of instances in this cluster is more than a threshold, a new classifier using these instances is trained and this classifier is added to the set of all trained classifiers. The loop stops when the number of samples in each cluster is less than a threshold.

For testing the participants, using the trained classifiers, all the instances of each participant are classified one by one using all the trained classifiers (line 2 of Algorithm 2). A subject would be classified in the target group if at least one of its instances is classified in the target group at least by one of the classifiers (lines 3 and 4 of Algorithm 2). Otherwise, if there is no instance classified among the target group, the participant is classified as the rest (lines 5 and 6).

Details of the implementations

The classifiers were implemented in Python using scikit-learn library.

WSI Classifiers:

We tested several common WSI classifiers, but we report only the result of SVM with RBF kernel and with no feature selection, which gives the best average accuracy. It should be noted that several feature selection approaches, like L1-SVM and backward elimination, were tested but they only reduced the accuracy. We used group 5-fold cross-validation for tuning hyper-parameters. Group K-fold means that all instances of each participant are placed in only one of the folds. This prevents having the same participant's instances in the train and validation folds simultaneously. In each

fold, there were two ASD and two TD participants. It should be mentioned that before applying the algorithms, we balanced the number of instances of the two groups using upsampling.

Two approaches were exploited to combine the decisions on different samples of a participant in the WSI approach. The first approach was majority pooling which classifies a participant as ASD if the number of instances classified as ASD was more than 50 percent of all instances. The second approach was threshold-based pooling which is similar to the first approach except that a threshold other than 50 percent is used.

SSI Classifiers:

Before applying the algorithm, we balanced the number of instances of the two groups by upsampling. The threshold for the minimum number of samples needed in a cluster, to be able to train a classifier is set to 10. It should be mentioned that agglomerative clustering and decision tree are the methods used for clustering and classification parts of Algorithm 1, respectively.

Training the SSI classifiers:

After running Algorithm 1 on our data, two exclusive clusters with enough instances, i.e. at least 10 instances in our study, were found. Then two classifiers were trained corresponding to each cluster. One of these exclusive clusters had 11 instances from 4 ASD participants (Table 1). These 11 instances consisted of 6 out of 9 instances of ASD1, 2 out of 4 instances of ASD10, 1 out of 2 instances of ASD8, and 2 out of 4 instances of ASD6. As explained in the algorithm, for each cluster, a decision tree classifier was trained using the ASD instances in the cluster versus all TD instances. Interestingly, only one feature was enough to discriminate instances in the cluster from all TD instances. Among those features that can discriminate the cluster's instances, we selected the Variance of Frame-wise Temporal Derivative (VFTD) of the 7th MFCC coefficient as the

feature which can discriminate more ASD participants from the set of all participants with a simple threshold. The classifier obtained by thresholding on this feature was the first classifier. This feature supports our expert's report regarding the higher variations in the cry of ASD children than TD children. From 10 ASD children, 8 of them can be discriminated using this feature. For each participant, the number of instances found by this classifier is shown in the 2nd column of Table 5.

Table 5. The number of instances of each participant in the training set that are classified as ASD using each trained SSI classifier.

ID	First SSI classifier	Second SSI classifier
ASD1	8	3
ASD2	1	2
ASD3	3	1
ASD4	10	9
ASD5	0	0
ASD6	1	3
ASD7	1	0
ASD8	1	2
ASD9	0	1
ASD10	2	4

After excluding the ASD samples from the first classifier, the second classifier was trained based on the second exclusive cluster. This cluster included all instances of participant ASD4. The only feature used for classifying this cluster was VFTD of the 6th SONE coefficient. SONE is a unit of loudness which is a subjective perception of sound pressure [52]. Having higher VFTD of the 6th SONE coefficient confirms the experiential knowledge of our experts mentioned before. Among all the ASD participant, eight have instances with VFTD of the 6th SONE higher than a threshold (Shown in the 3rd column of Table 5). The results of classification based on these two features are depicted in Fig 3. As it is mentioned in the proposed method section, the participants with at least one instance classified into this cluster would be considered as a participant with ASD.

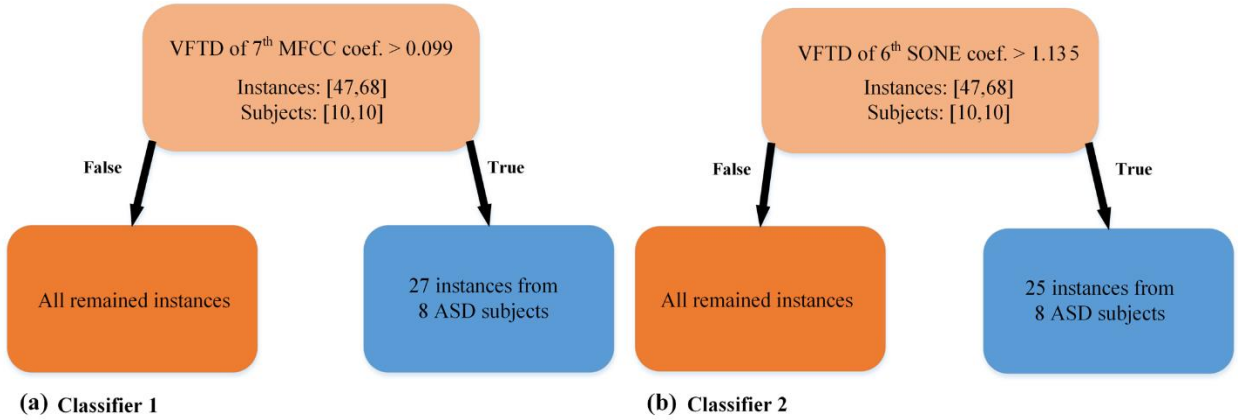


Fig 3 Two classifiers trained on the two exclusive clusters found during the SSI classifier training phase. (a) The Variance of Frame-wise Temporal Derivative (VFTD) of the 7th MFCC coefficient separates 27 instances of 8 ASD subjects from all TD instances of the training set. (b) VFTD of the 6th SONE coefficient separates 17 instances of 7 ASD participants from all TD instances of the training set.

Results

In this part, the performance of our proposed SSI classifier against a common WSI classifier is evaluated on our test set of ASD and TD participants. Each participant has multiple instances which are cleaned using the criteria explained in the data collection and preprocessing section. The participants who had at least one accepted instance were used in the training and testing phases, which are shown in Tables 1 and 2.

The output of the SSI approach was two classifiers, each of them works by thresholding on a feature. The number of instances of ASD participants in the training set, correctly detected by the first and the second classifiers, are shown in the second and third columns of Table 5, respectively.

On the other hand, the best-resulting classifier for the WSI approach was Radial Basis Function-Support Vector Machine (RBF-SVM) [53].

The classification results on the test set for different classifiers are shown in Table 6. The portion of each participant's instances correctly classified by each classifier is written in percentage under the name of the classifier. The decision made by the WSI and SSI classifiers for each participant is shown by ASD or TD. To classify each subject using the WSI classifier, the Majority Pooling (MP) and the Best-chance threshold Pooling (BP) approaches are used. BP is a threshold-based pooling with the threshold giving the best accuracy on the test set for male participants. For the boys, MP has specificity, sensitivity, and precision equal to 100%, 35.71%, and 67.85%, respectively. On the other hand, BP leads to specificity, sensitivity, and precision equal to 85.71%, 71.42%, and 78.57%, respectively. The threshold for BP was set to 20% that means if 20% of instances of a participant were classified as ASD instance, the participant was classified as ASD. The results of the percentage of instances correctly classified by the two classifiers in the SSI approach are shown as C_1 (the first SSI classifier) and C_2 (the second SSI classifier) in Table 6. The aggregated result of C_1 and C_2 makes the final decision of the SSI classifier which is shown in the decision column under the SSI classification section. The achieved specificity, sensitivity, and precision using the proposed method for the boys are 100%, 85.71%, and 92.85%, respectively.

Table 6. The results of classifiers on the instances of each participant in the test set.

	TD children							Children with ASD						
	ID	Portion of instances classified as TD in percentage and the decision						ID	Portion of instances classified as ASD in percentage and the decision					
		WSI classification			SSI classification				WSI classification			SSI classification		
		SVM	Dec.		C_1	C_2	Dec.		SVM	Dec.		C_1	C_2	Dec.
			MP	BP						MP	BP			
Boys	TD11	100	TD	TD	100	100	TD	ASD11	50	A	A	17	50	ASD
	TD12	100	TD	TD	100	100	TD	ASD12	33	TD	ASD	11	28	ASD
	TD13	100	TD	TD	100	100	TD	ASD13	33	TD	ASD	33	0	ASD

	TD14	100	TD	TD	100	100	TD	ASD14	20	TD	ASD	20	20	ASD
	TD15	100	TD	TD	100	100	TD	ASD15	0	TD	TD	0	40	ASD
	TD16	100	TD	TD	100	100	TD	ASD16	50	ASD	ASD	100	0	ASD
	TD17	100	TD	TD	100	100	TD	ASD17	0	TD	TD	0	100	ASD
	TD18	83	TD	TD	100	100	TD	ASD18	50	ASD	ASD	50	50	ASD
	TD19	100	TD	TD	100	100	TD	ASD19	0	TD	TD	0	0	TD
	TD20	80	TD	ASD	100	100	TD	ASD20	42	TD	ASD	42	16	ASD
	TD21	100	TD	TD	100	100	TD	ASD21	100	ASD	ASD	0	0	TD
	TD22	100	TD	TD	100	100	TD	ASD22	0	TD	TD	0	50	ASD
	TD23	75	TD	ASD	100	100	TD	ASD23	33	TD	ASD	33	17	ASD
	TD24	92	TD	TD	100	100	TD	ASD24	86	ASD	ASD	86	86	ASD
	Acc. %		100	85.71			100			35.71	71.42			85.71
Girls	TD25	100	TD	TD	100	100	TD	ASD25	42	TD	ASD	17	0	ASD
	TD26	100	TD	TD	100	100	TD	ASD26	60	ASD	ASD	60	20	ASD
	TD27	100	TD	TD	100	100	TD	ASD27	50	ASD	ASD	0	0	TD
	TD28	100	TD	TD	100	100	TD	ASD28	100	ASD	ASD	0	50	ASD
	TD29	100	TD	TD	100	100	TD	ASD29	62	ASD	ASD	50	50	ASD
	TD30	67	TD	ASD	100	100	TD	ASD30	100	ASD	ASD	50	50	ASD
	TD31	100	TD	TD	100	100	TD	ASD31	0	TD	TD	0	0	TD
	Acc. %		100	85.71			100			71.42	85.71			71.42

Each classifier result on a subject's instances are reported in percentage.

Dec**.,** Decision; MP, Majority Pooling; BC, Best-chance threshold Pooling; C1, Classifier1; C2, Classifier2; Acc**.,** Accuracy.

To further show the applicability of the proposed approach to girls, we applied the boys' trained classifiers on the test set of the girls. The results are shown in the last row of Table 6 which show that the MP approach has specificity, sensitivity, and precision equal to 100%, 71.42%, and 85.71%, respectively. Furthermore, the BP approach gives specificity, sensitivity, and precision all equal to 85.71%, respectively. The results of the proposed SSI classifier is 100% specificity, 71.42% sensitivity, and 85.71% precision.

A two-dimensional scatter plot of the two features, used in C_1 and C_2 classifiers, are shown in Fig 4. As it can be seen in this figure, the instances of a participant with ASD are scattered in the area containing instances of both TD and ASD participants. Nevertheless, there are instances for this participant uniquely distinguishable using the selected two features.

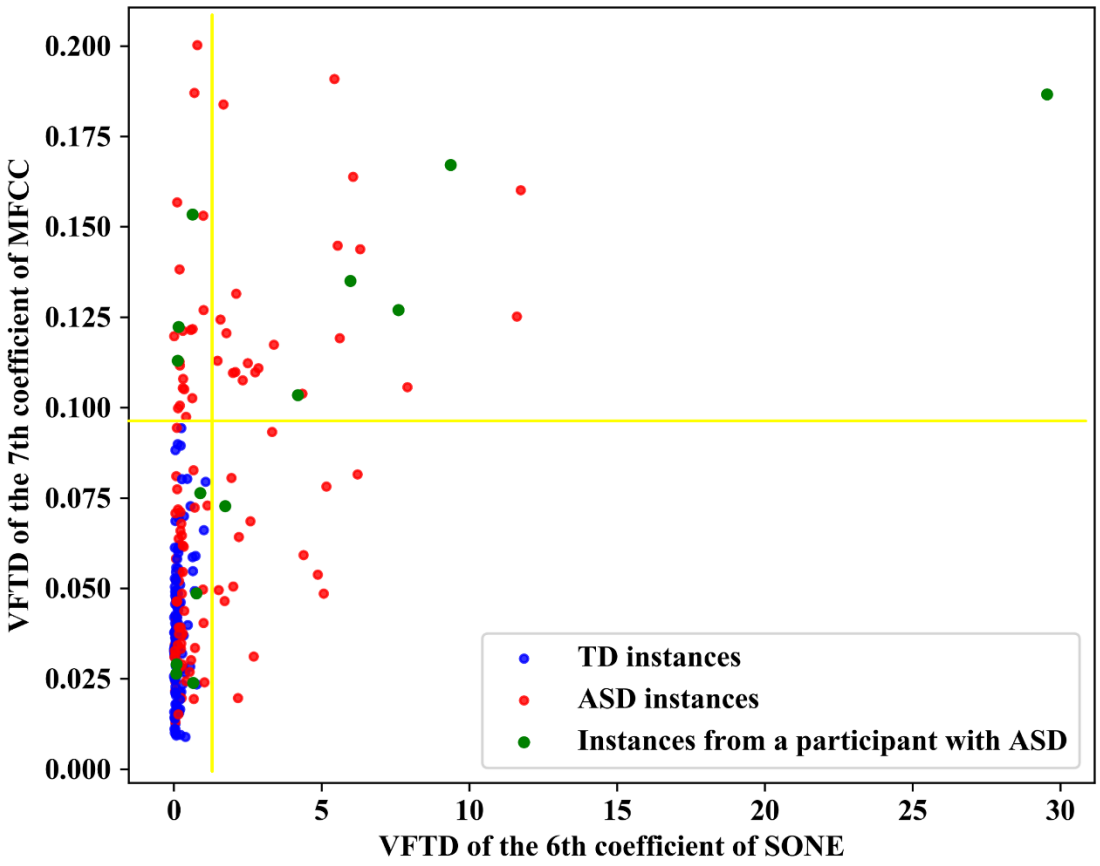


Fig 4. Instances of several ASD and TD participants scattered in the space of two features given by the proposed SSI method. The instances of a chosen ASD participant are illustrated in green to show that a participant may have instances in the area common with TD instances besides those two areas separated by the selected thresholds as ASD. The mentioned ASD participant (with green instances) is tagged as ASD due to having at least one instance with the greater value than at least one of the thresholds on the two features.

We compared the results of our proposed method with that of the only method available in the literature which was trained using only cry features [36] based on our data. The results (Table 7) show the superiority of our method compared to the previously proposed method.

Table 7. Comparison of the results on the test set using the two methods; SSI approach and a baseline approach.

		Sensitivity	Specificity	Precision
Boys	SSI	85.71%	100%	92.85%
	Baseline	50.58%	81%	65%
Girls	SSI	71.42%	100%	85.71%
	Baseline	21%	86.48%	53%

Investigating the trained classifier on participants under 18 months

The SSI classifier which was trained using the training set in Table 1 was also tested on the data of children younger than 18 months. From 57 participants under 18 months, two boys (Child1 and Child2 in Table 3) were classified as ASD by the mentioned trained classifier. These participants were referred to our experts for diagnosis. These two were suspected to have neurodevelopmental problems. All other boys were classified as TD. However, among them, Child3 was diagnosed with ASD at age 2. Also, Child4 showed symptoms of having ADHD and sensory processing disorder at age 3. Three other children had symptoms which suggested that they are not TD children. Two of the girls who were 18 months old were classified as ASD using the trained classifier. The other girls were classified as TD. The results of testing the trained SSI classifier on this data set are summarized in Table 8.

Table 8. Classification of the participants under 18 months using our trained SSI classifier.

	Boys			Girls		
	ASD	TD	Others[a]	ASD	TD	Others[a]
Classified as ASD	0	0	2	0	1	0
Classified as TD	1	22	4	0	27	0

[a] Other developmental or mental disorders

Discussion and conclusion

In this paper, we presented a novel cry-based screening method to distinguish between children with autism and typically developing children. In the proposed method, groups of children with autism who have specific features in their cry voices can be determined. This method is based on a new classification approach called SubSet Instance (SSI) classifier. A nice property of the proposed SSI classifier, in the case of voice-based autism screening, is its high specificity such that a normal child can be detected with no error. We applied the proposed method on a group of participants consisted of 24 boys with ASD between 20 and 53 months of age and 24 TD boys between 18 and 51 months of age. The two features, found in this study, were used to train a classifier on 10 boys with ASD and 10 TD boys. Then, the classifier was used to distinguish 14 boys with ASD from 14 TD boys, reaching 92.8% accuracy. Due to the fact that girls are less likely to have autism and consequently, it is harder to collect enough data from girls than boys, the number of girls with ASD was not enough to train a separate classifier for this gender. It should be noted that we tested the trained system on 7 girls with ASD and 7 TD girls. It was shown that the trained classifier can screen girls with 7% lower accuracy than boys of the test set. In other words, it seems that gender matters and it should be considered in the training of the system. In testing the data from participants under 18 months, one TD girl was classified as ASD which was not the case for any TD children of the male counterparts. This result also confirms the mentioned note about the gender effect. However, in our future work, we would try to collect more data of girls to be able to train a system to accurately screen girls. In the future, we would also try to train a single classifier for boys and girls to determine whether it can be used for both of them.

It should be mentioned that our training and test data were completely separate making the trained model more general. The features found in this study were applicable in the range of ages of our

participants from 18 to 53 months. This is in contrast to other approaches that either used a dataset of children with a specific age [28, 30] or used age information for classification [29]. Due to the age invariant features found in this study, it can be claimed that there are markers in the voice of children with ASD that are sustained at least in a range of ages.

The two discriminative features found in this study were a coefficient of MFCC and a SONE coefficient. MFCC and SONE are related to the power spectrum of a speech signal. SONE measures loudness in specific Bark bands [50]. On the other hand, MFCC, which is the inverse DFT of log-spectrum in the Mel scale, is related to the timbre of the voice [54]. Therefore, MFCC and SONE can be interpreted to be related to the timbre and loudness of a tone. Furthermore, based on the feedback from our experts, there is unpredictability in the crying voice of children with autism which is not true for TD children. Consequently, we used the variance of temporal difference as a feature suitable for screening children with autism. This is due to the fact that if a signal is constant or changed linearly over time, the variance of temporal difference is zero. Therefore, the variance of temporal difference can be seen as the amount of ambiguity or unpredictability of a sound. On the other hand, the heightened variability in the two features, found in this study, for children with ASD is significant due to the reports from other studies [17, 55] which shows increased biological signals variability in children with ASD and infants at high risk for autism in comparison to TD children. These features are statistical features of the cry instances that hold constant, at least, across an age range studied in this research.

To the best of our knowledge, [29] and [30] were the only studies on screening children with autism using voice features on children younger than 2 years old. Our proposed method has higher precision than these two, i.e. 6% more than [29] and 17% more than [30], using only cry features. The use of cry features as suitable biomarkers for autism screening matches the claims in [33].

In our study only children with ASD and TD children were tested. Other developmental disorders or health issues were not tested to see how children with those disorders would be classified using the proposed method which can decrease the specificity of 100%. However, this approach is proposed to be used as a screening tool and final diagnosis should be done under experts' supervision. So, this approach can be applied as a general screener of autism spectrum disorder.

The trained classifier was also tested on 57 participants between 10 to 18 months of age. The classifier screened two boys from the rest, i.e. Child1 and Child2 (Table 3). Child1 showed evidences of genetic disease and diagnosed with developmental delay and Child2 received UNDD classification by our experts. This suggests that a) the system can be used for children under 2 years of age, and b) it may be able to distinguish other neurodevelopmental disorders. On the other hand, there were 5 boys, i.e. Child3 to Child7 (Table 3), who had no evidence of mental or developmental disorders at the time of their recording. At the same time, our approach did not distinguish them as children with ASD either. However, when they got older than 3 years, they showed symptoms of neurodevelopmental disorders. From these children, Child3 and Child4's voices, collected after receiving the diagnosis, were classified as children with ASD using our approach. Unfortunately, Child5, Child6, and Child7 did not cooperate to be evaluated by an expert to validate the results of our expert-selected questionnaire. Furthermore, they did not cooperate to send us their children's recent cry voices.

The result of studying these 57 children under age of 18 months may suggest that: a) there could be symptoms in the crying voices of children with neurodevelopmental disorders under 18 months (Child1 and Child2), b) the approach may not be able to screen a participant with neurodevelopmental disorders under the age of 18 months with the possibility of: 1) the participant was among those children with neurodevelopmental disorder who do not have our proposed

specific features in their crying voices, 2) the participant's recorded cry samples did not include our specific features, and/or 3) Neurodevelopmental disorders and their features had not been developed in the child at the time of initial recording. The reason behind not classifying Child3 and Child4, as children with ASD under the age of 18, could be b.2 or b.3. To clearly determine the reason behind this phenomena, further investigation is needed.

We believe that this approach can be used to perform early autism screening under 18 months of age. Thus, in the future, we need to collect data and test the approach on more data of children under 18 months to validate these results with more confidence.

We have to further check the proposed approach and the extracted features on other neurodevelopmental disorders, such as ADHD, to evaluate the capability of the approach to distinguish the children with these disorders from TD children.

Furthermore, without comparing the cries of children with ASD to those without ASD but another disorder, we don't really know if these findings are specific to autism or to general atypical brain developments. Thus, we should collect cries of children with other neurodevelopmental disorders and compare voices of children with ASD to voices of other neurodevelopmental disorders to see if these features would be able to separate them or not.

It is shown that crying consists of intricate motor activities [56]. On the other hand, it is shown that children with ASD have problems in the motor domain and in coordination of their motor capabilities with other modalities [57]. Consequently, it is possible that the extracted features in the crying voices of children with ASD come from this deficiency/problem in the motor domain which needs further investigations.

Finally, automating the preprocessing part is a technical issue that should be handled if the cry voice-based screening is planned to be fully automated. This would be important since such a screening system can be deployed in systems such as Amazon Alexa [58] to automatically screen problematic cry voices.

Acknowledgements

We would like to thank the Center for Treatment of Autism Disorder (CTAD) and its members for supporting this study. Also, we like to thank all the families who helped to collect the cry voices of their children. The authors also like to thank Prof. H. Sameti from Sharif University of Technology for his great feedbacks on the data collection and voice processing.

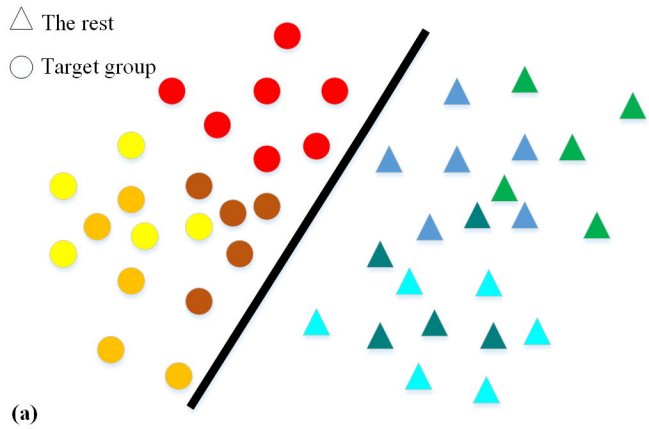
References:

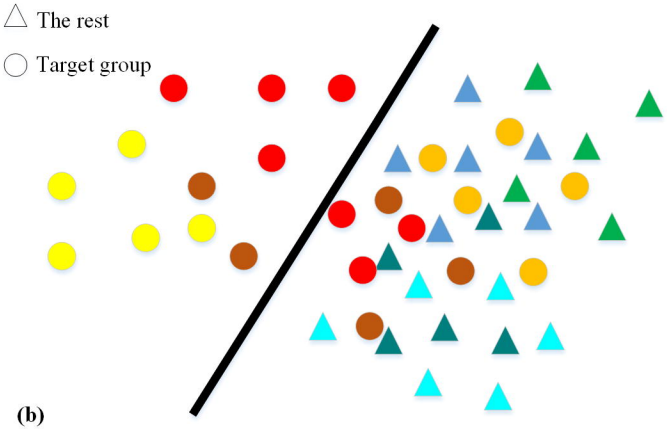
1. American Psychiatric Association. Diagnostic and statistical manual of mental disorders (DSM-5®). American Psychiatric Pub; 2013.
2. Chen JL, Sung C, Pi S. Vocational rehabilitation service patterns and outcomes for individuals with autism of different ages. J Autism Dev Disord. 2015;45(9):3015-29.
3. Fakhoury M. Autistic spectrum disorders: A review of clinical features, theories and diagnosis. Int J Dev Neurosci. 2015;43:70-7.
4. Constantino JN, Charman T. Diagnosis of autism spectrum disorder: reconciling the syndrome, its diverse origins, and variation in expression. Lancet Neurol. 2016;15(3):279-91.
5. Calderoni S, Billeci L, Narzisi A, Brambilla P, Retico A, Muratori F. Rehabilitative interventions and brain plasticity in autism spectrum disorders: focus on MRI-based studies. Front Neurosci. 2016;10:139.
6. Brentani H, Paula CSd, Bordini D, Rolim D, Sato F, Portolese J, et al. Autism spectrum disorders: an overview on diagnosis and treatment. Braz J Psychiatry. 2013;35:S62-S72.
7. Mandell DS, Novak MM, Zubritsky CD. Factors associated with age of diagnosis among children with autism spectrum disorders. Pediatrics. 2005;116(6):1480-6.
8. Thabtah F, Peebles D. A new machine learning model based on induction of rules for autism detection. Health Inform J. 2019. Doi: 10.1177/1460458218824711.
9. Zachor DA, Itzchak EB. Treatment approach, autism severity and intervention outcomes in young children. Res Autism Spectr Disord. 2010;4(3):425-32.
10. Thabtah F, Peebles D. Early Autism Screening: A Comprehensive Review. Int J Environ Res Public Health. 2019;16(18):3502.
11. Rutter M, Le Couteur A, Lord C. ADI-R: Autism Diagnostic Interview-Revised. Los Angeles, CA: Western Psychological Services; 2003.
12. Lord C., Risi S., Lambrecht L., Cook E. H. Jr., Leventhal B. L., Di Lavore, et al. The autism diagnostic observation schedule-generic: a standard measure of social and communication deficits associated with the spectrum of autism. J Autism Dev Disord. 2000; 30(3), 205–223.
13. Levy S, Duda M, Haber N, Wall DP. Sparsifying machine learning models identify stable subsets of predictive features for behavioral detection of autism. Mol Autism. 2017;8(1):65. Doi: 10.1186/s13229-017-0180-6.
14. Küpper C, Stroth S, Wolff N, Hauck F, Kliewer N, Schad-Hansjosten T, et al. Identifying predictive features of autism spectrum disorders in a clinical sample of adolescents and adults using machine learning. Sci Rep. 2020;10(1):4805. Doi: 10.1038/s41598-020-61607-w.
15. Abbas H, Garberson F, Liu-Mayo S, Glover E, Wall DP. Multi-modular AI Approach to Streamline Autism Diagnosis in Young Children. Scientific Reports. 2020;10(1):5014. Doi: 10.1038/s41598-020-61213-w.
16. Emerson RW, Adams C, Nishino T, Hazlett HC, Wolff JJ, Zwaigenbaum L, et al. Functional neuroimaging of high-risk 6-month-old infants predicts a diagnosis of autism at 24 months of age. Sci Transl Med. 2017;9(393):eaag2882.
17. Denisova K, Zhao G. Inflexible neurobiological signatures precede atypical development in infants at high risk for autism. Sci Rep. 2017;7(1):11285. Doi: 10.1038/s41598-017-09028-0.
18. Bosl WJ, Tager-Flusberg H, Nelson CA. EEG analytics for early detection of autism spectrum disorder: a data-driven approach. Sci Rep. 2018;8(1):6828.
19. Momeni N, Bergquist J, Brudin L, Behnia F, Sivberg B, Joghataei M, et al. A novel blood-based biomarker for detection of autism spectrum disorders. Transl Psychiatry. 2012;2(3):e91.

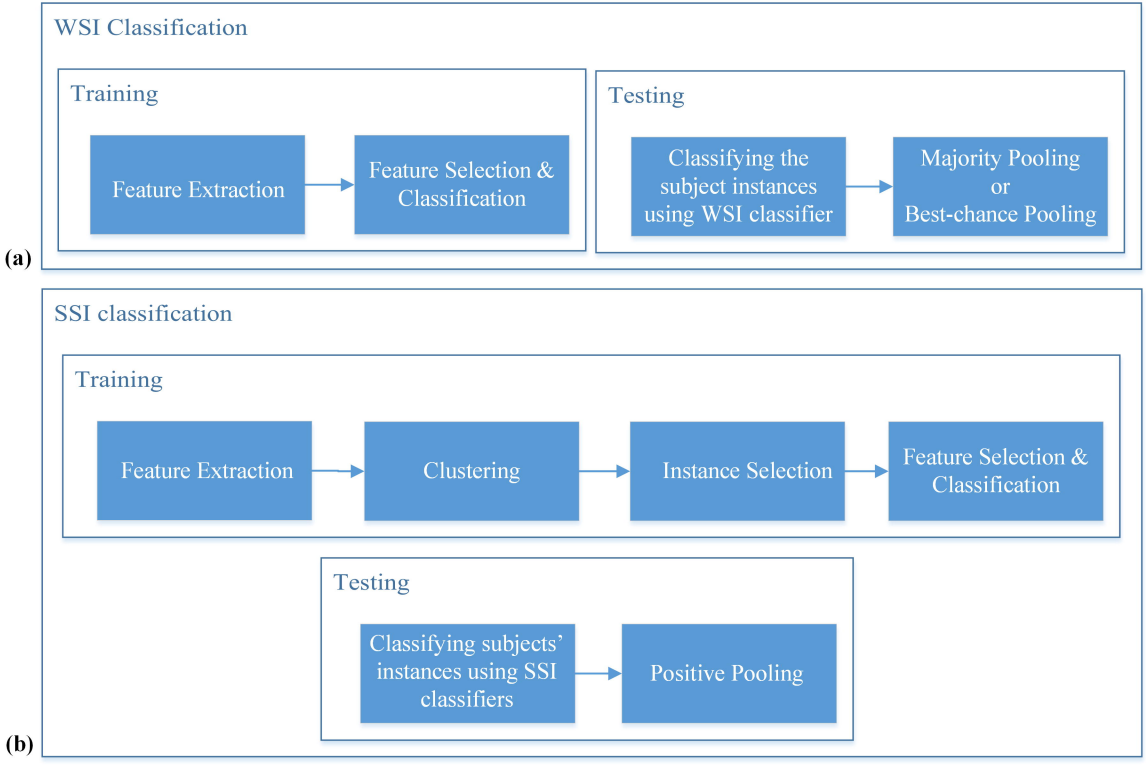
20. Glatt SJ, Tsuang MT, Winn M, Chandler SD, Collins M, Lopez L, et al. Blood-based gene expression signatures of infants and toddlers with autism. J Am Acad Child Adolesc Psychiatry. 2012;51(9):934-944.e2.
21. Croen LA, Braunschweig D, Haapanen L, Yoshida CK, Fireman B, Grether JK, et al. Maternal mid-pregnancy autoantibodies to fetal brain protein: the early markers for autism study. Biol Psychiatry. 2008;64(7):583-8.
22. Greene DJ, Black KJ, Schlaggar BL. Considerations for MRI study design and implementation in pediatric and clinical populations. Dev Cogn Neurosci. 2016;18:101-112. Doi: 10.1016/j.dcn.2015.12.005.
23. Webb SJ, Bernier R, Henderson HA, Johnson MH, Jones EJ, Lerner MD, et al. Guidelines and best practices for electrophysiological data collection, analysis and reporting in autism. J Autism Dev Disord. 2015;45(2):425-243.
24. Engelhardt LE, Roe MA, Juranek J, DeMaster D, Harden KP, Tucker-Drob EM, et al. Children's head motion during fMRI tasks is heritable and stable over time. Dev Cogn Neurosci. 2017;25:58-68. Doi: 10.1016/j.dcn.2017.01.011.
25. Denisova K. Age attenuates noise and increases symmetry of head movements during sleep resting-state fMRI in healthy neonates, infants, and toddlers. Infant Behav Dev. 2019;57:101317. Doi: 10.1016/j.infbeh.2019.03.008.
26. Brisson J, Martel K, Serres J, Sirois S, Adrien JL. Acoustic analysis of oral productions of infants later diagnosed with autism and their mother. Infant Ment Health J. 2014;35(3):285-95.
27. Nakai Y, Takiguchi T, Matsui G, Yamaoka N, Takada S. Detecting abnormal word utterances in children with autism spectrum disorders: machine-learning-based voice analysis versus speech therapists. Percept Mot Skills. 2017;124(5):961-973.
28. Santos JF, Brosh N, Falk TH, Zwaigenbaum L, Bryson SE, Roberts W, et al. Very early detection of autism spectrum disorders based on acoustic analysis of pre-verbal vocalizations of 18-month old toddlers. In: Proceedings of the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing; 2013; Vancouver, BC, Canada: IEEE. 2013. Doi: 10.1109/ICASSP.2013.6639134.
29. Oller D, Niyogi P, Gray S, Richards J, Gilkerson J, Xu D, et al. Automated vocal analysis of naturalistic recordings from children with autism, language delay, and typical development. Proc Natl Acad Sci. 2010;107(30):13354-9.
30. Pokorny FB, Schuller BW, Marschik PB, Brueckner R, Nyström P, Cummins N, et al. Earlier Identification of Children with Autism Spectrum Disorder: An Automatic Vocalisation-Based Approach. In: Proceedings of the INTERSPEECH 2017; 2017; Stockholm, Sweden: ISCA. 2017. Doi: 10.21437/Interspeech.2017-1007.
31. Little MA, Varoquaux G, Saeb S, Lonini L, Jayaraman A, Mohr DC, et al. Using and understanding cross-validation strategies. Perspectives on Saeb et al. Gigascience. 2017;6(5):1-6. Doi: 10.1093/gigascience/gix020.
32. Eyben F, Scherer KR, Schuller BW, Sundberg J, André E, Busso C, et al. The Geneva minimalistic acoustic parameter set (GeMAPS) for voice research and affective computing. IEEE Trans Affect Comput. 2015;7(2):190-202.
33. Esposito G, Hiroi N, Scattoni ML. Cry, Baby, Cry: Expression of Distress As a Biomarker and Modulator in Autism Spectrum Disorder. Int J Neuropsychopharmacol. 2017;20(6):498-503.
34. Sheinkopf SJ, Iverson JM, Rinaldi ML, Lester BM. Atypical Cry Acoustics in 6-Month-Old Infants at Risk for Autism Spectrum Disorder. Autism Res. 2012;5(5):331-9.

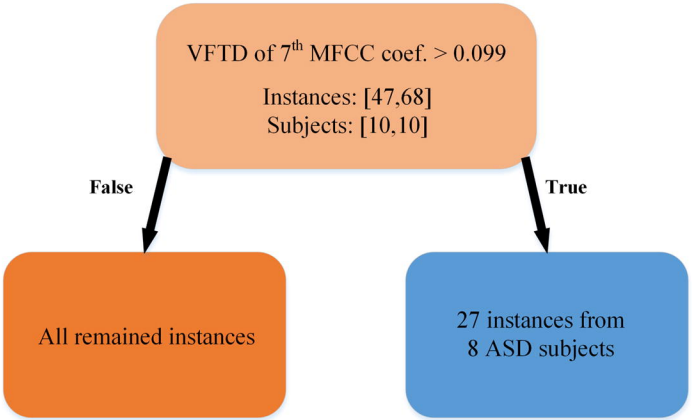
35. Orlandi S, Manfredi C, Bocchi L, Scattoni ML, editors. Automatic newborn cry analysis: a non-invasive tool to help autism early diagnosis. In: Proceedings of the 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society; 2012; San Diego, CA, USA: IEEE. 2012. Doi: 10.1109/EMBC.2012.6346583
36. Motlagh SHRE, Moradi H, Pouretemad H, editors. Using general sound descriptors for early autism detection: 2013 9th Asian Control Conference (ASCC) Control; 2013; Istanbul, Turkey: IEEE. 2013. Doi: 10.1109/ASCC.2013.6606386
37. Barrow WJ, Jaworski M, Accardo PJ. Persistent toe walking in autism. J Child Neurol. 2011;26(5):619-21.
38. Paul R, Norbury CF. Language disorders from infancy through adolescence: Elsevier; 2012.
39. Jalilevand N, Ebrahimipour M. Pronoun acquisition in Farsi-speaking children from 12 to 36 months. J Child Lang Acquis Dev. 2013;1(1):1-9.
40. Goldstein S, Ozonoff S. Assessment of autism spectrum disorder: Guilford Publications; 2018.
41. Lund NJ, Duchan JF. Assessing children's language in naturalistic contexts: Prentice Hall; 1993.
42. Gilliam JE. Gilliam autism rating scale: GARS 2: Pro-ed; 2006.
43. Berk L. Development through the lifespan: Pearson Education India; 2010.
44. Three ZT. Diagnostic classification of mental health and developmental disorders of infancy and early childhood: Revised edition (DC: 0-3R). Washington, DC: Zero To Three Press; 2005.
45. Molau S, Pitz M, Schluter R, Ney H. Computing Mel-frequency cepstral coefficients on the power spectrum. In: Proceedings of the 2001 IEEE International Conference on Acoustics, Speech, and Signal Processing Proceedings (Cat No01CH37221); 2001; Salt Lake City, UT, USA: IEEE; 2001. Doi: 10.1109/ICASSP.2001.940770
46. Esposito G, Venuti P. Developmental changes in the fundamental frequency (f0) of infants' cries: a study of children with Autism Spectrum Disorder. Early Child Dev Care. 2010;180(8):1093-102.
47. Marchi E, Schuller B, Baron-Cohen S, Golan O, Bölte S, Arora P, et al. Typicality and emotion in the voice of children with autism spectrum condition: Evidence across three languages. In: Proceedings of the INTERSPEECH 2015; 2015; Dresden, Germany: ISCA. 2015. p. 115-119. Available from: https://www.isca-speech.org.
48. Belalcázar-Bolaños E.A., Orozco-Arroyave J.R., Vargas-Bonilla J.F., Haderlein T., Nöth E. Glottal Flow Patterns Analyses for Parkinson's Disease Detection: Acoustic and Nonlinear Approaches. In: Sojka P., Horák A., Kopeček I., Pala K., editors. Text, Speech, and Dialogue: Proceedings of the 19th International Conference on Text, Speech, and Dialogue; 2016 Sep 12-16; Brno , Czech Republic. Cham: Springer; 2016. Doi : 10.1007/978-3-319-45510-5_46.
49. Rabiner LR, Schafer RW. Introduction to digital speech processing. Found and trends in signal process. 2007;1(1):1-194.
50. Peeters G. A large set of audio features for sound description (similarity and classification) in the CUIDADO project. CUIDADO IST Proj Rep. 2004;54(0):1-25.
51. Bone D, Lee C-C, Black MP, Williams ME, Lee S, Levitt P, et al. The psychologist as an interlocutor in autism spectrum diorder assessment: Insights from a study of spontaneous prosody. J Speech Lang Hear Res. 2014;57(4):1162-77.
52. Hänsler E, Schmidt G. Speech and audio processing in adverse environments: Springer Science & Business Media; 2008.
53. Theodoridis S, Koutroumbas K. Pattern recognition: Elsevier; 2003.

54. Li TL, Chan AB. Genre classification and the invariance of MFCC features to key and tempo. In: Lee KT., Tsai WH., Liao HY.M., Chen T., Hsieh JW., Tseng CC., editors. Advances in Multimedia Modeling: Proceedings of the 17th International MultiMedia Modeling Conference; 2011 Jan 5-7; Taipei, Taiwan. Berlin, Heidelberg: Springer; 2011. Doi: 10.1007/978-3-642-17832-0_30.
55. Takahashi T, Yoshimura Y, Hiraishi H, Hasegawa C, Munesue T, Higashida H, et al. Enhanced brain signal variability in children with autism spectrum disorder during early childhood. Hum Brain Mapp. 2016;37(3):1038-50.
56. Lester BM, Boukydis CZ. Infant crying: Theoretical and research perspectives: Springer; 1985.
57. MacDonald M, Lord C, Ulrich D. The relationship of motor skills and adaptive behavior skills in young children with autism spectrum disorders. Res Autism Spectr Disord. 2013;7(11):1383-90.
58. Hoy MB. Alexa, Siri, Cortana, and More: An Introduction to Voice Assistants. Medical Reference Services Quarterly. 2018;37(1):81-8. Doi: 10.1080/02763869.2018.1404391.

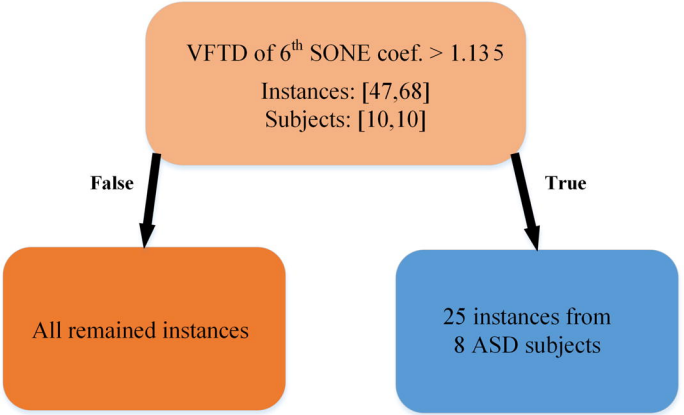

The rest
Target group
(a)


The rest
Target group
(b)

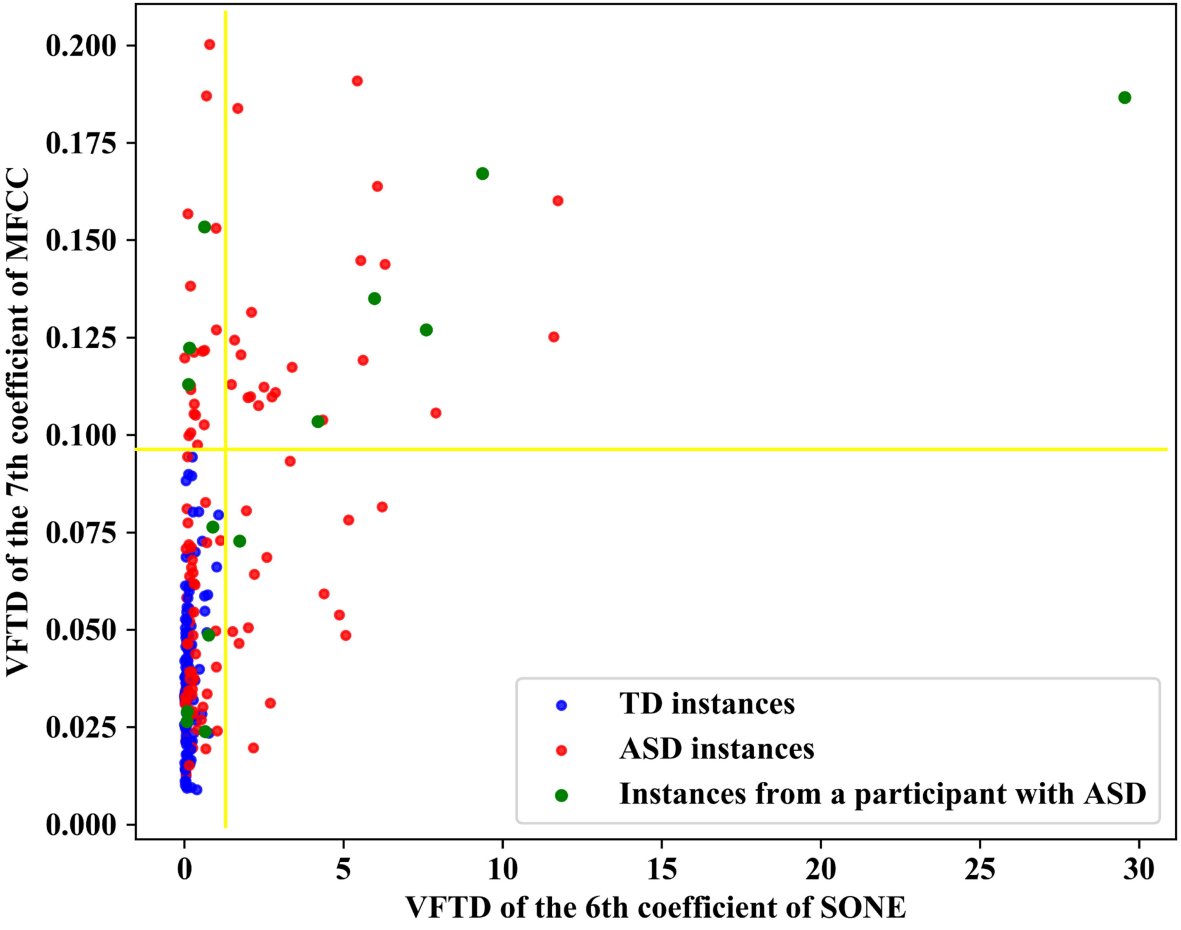

WSI Classification
Training
Feature Extraction
Feature Selection & Classification
Testing
Classifying the subject instances using WSI classifier
Majority Pooling or Best-chance Pooling
(a)
SSI classification
Training
Feature Extraction
Clustering
Instance Selection
Feature Selection & Classification
Testing
Classifying subjects' instances using SSI classifiers
Positive Pooling
(b)



(a) Classifier 1



(b) Classifier 2


0.200
0.175
0.150
0.125
0.100
0.075
0.050
0.025
0.000
VFTD of the 7th coefficient of MFCC
0
5
10
15
20
25
30
VFTD of the 6th coefficient of SONE
TD instances
ASD instances
Instances from a participant with ASD